



Allan, N., Thomas, L., Hart, J., Freeman, C., & Mohn, C. (2018). Calcite–magnesite solid solutions: using genetic algorithms to understand nonideality. *Physics and Chemistry of Minerals*.
<https://doi.org/10.1007/s00269-018-0997-3>

Peer reviewed version

License (if available):
Unspecified

Link to published version (if available):
[10.1007/s00269-018-0997-3](https://doi.org/10.1007/s00269-018-0997-3)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer at <https://link.springer.com/article/10.1007/s00269-018-0997-3> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

1 **Calcite-magnesite solid solutions: using genetic algorithms to understand**

2 **non-ideality**

3 N.L. Allan, and L. Thomas

4 *School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK*

5 J.N. Hart

6 *School of Materials Science and Engineering, UNSW Australia,*

7 *UNSW Sydney NSW 2052, Australia*

8 C.L. Freeman

9 *Materials Science and Engineering Department, Kroto Research Institute, University of*

10 *Sheffield, Broad Lane, Sheffield, S3 7HQ, UK*

11 C.E. Mohn

12 *Centre for Earth Evolution and Dynamics, University of Oslo, P.O. Box 1048 Blindern,*

13 *N0316 Oslo, Norway.*

14 **ABSTRACT**

15 We show how a genetic algorithm (GA) generates efficiently the energy landscape of the
16 equimolar calcite-magnesite ($\text{CaCO}_3 - \text{MgCO}_3$) solid solution. Starting from a random
17 configuration of cations and a suitable supercell, the lowest-energy form of ordered dolomite
18 emerges rapidly. Practical implementation and operation of the GA are discussed in detail.
19 The method can also generate both low-lying and high-lying excited states. Detailed analysis
20 of the energy-minimised structures of the different configurations reveals that low energies
21 are associated with reduction of strain associated with rotation of the carbonate groups, a
22 mechanism possible only when a carbonate layer lies between a layer of just Ca and a layer
23 of just Mg. Such strain relief is not possible in the equimolar MgO-CaO solid solution despite
24 the similarity of the crystal structures of these binary oxides to calcite-magnesite, and so the
25 enthalpy of mixing is very high. Implications for thermodynamic configurational averaging

26 over the minima in the energy landscape are briefly considered. Overall, the genetic
27 algorithm is shown to be a powerful tool in probing non-ideality in solid solutions and
28 revealing the ordering patterns that give rise to such behaviour.

29 INTRODUCTION

30 Phases with constant composition are rare in geology. Most natural mineral groups
31 exist over a range of chemical composition and solid solutions play a major role in
32 determining mineral stability and chemical and physical behaviour. Ideality or non-ideality of
33 such solutions and in turn the resulting tendency to mix or unmix often depends on the
34 atomic ordering. For example, the enthalpy of mixing of disordered dolomite is positive, but
35 negative for ordered dolomite. Similar behaviour is found in the diopside-jadeite solid
36 solution - it is non-ideal, yet an intermediate ordered phase forms on cooling. Non-ideality is
37 fundamental also to the interpretation of any processes involving partitioning between phases.
38 Solid solutions continue to pose considerable challenges for computation, as does non-
39 ideality in particular. In this paper we report the use of genetic algorithms to *predict* the
40 existence of an ordered phase ab initio and discuss why this phase is more stable than the
41 disordered form.

42

43 We have developed a number of computational methods for the study of solid solutions and
44 grossly non-stoichiometric compounds. Any technique must be able to sample many
45 arrangements of the atoms, allowing for the exchange of ions located at different positions. It
46 is also crucial to allow for the relaxation of the local environment of each ion, i.e., local
47 structural movements which can reduce considerably the energy associated with ion
48 exchange. Local effects due to ion association or clustering must not be averaged out. We
49 have previously used both basin-sampling approaches such as configurational Boltzmann
50 averaging (Purton et al., 1998; Allan et al., 2001; Todorov et al., 2004; Mohn et al., 2005) and
51 exchange Monte Carlo methods (Purton et al., 1998, Todorov et al., 2004).

52

This paper is primarily concerned with basin-sampling and exploring the energy landscape of a strongly non-ideal system, $\text{MgCO}_3\text{-CaCO}_3$. Such marked non-ideality poses a number of difficult problems. Unlike the energy landscapes of the binary oxide mixtures considered in our previous work, (Purton et al., 1998) where configurations with distinct cation arrangements were generated randomly, the energy landscape for $\text{MgCO}_3\text{-CaCO}_3$ is very different. It contains a small number of deep minima, some of which correspond to the formation of ordered dolomite at 50% Mg, 50% Ca. Generating cation arrangements at random is almost bound to fail to discover such deep minima with very small weights (degeneracies). Since only a few minima are thermally accessible, simply generating a random selection of starting configurations will not probe sufficiently the low-energy parts of the landscape, and so any averaged thermodynamic property is likely to be highly inaccurate. Monte-Carlo techniques are also likely to fail to locate such minima due to “basin trapping”; the large mismatch in ionic radii, (Shannon, 1976) between Ca^{2+} (1.00 Å) and Mg^{2+} (0.72 Å) is such that changes in local environments associated with any exchange are large and so the acceptance rate of exchanges is very low even at high temperatures.

The development of tools for locating low-lying minima for such situations is of considerable importance. Ordering patterns may be extremely complex and somewhat counter-intuitive, such as in garnet solid solution, where third-, fourth- and even fifth-nearest neighbour cation orderings are energetically more important than first and second neighbours (van Westeren et al., 2003). Before tackling systems where ordering is not fully established, in this paper we take a well understood system and show how a genetic algorithm (GA) together with energy minimisation can be used to find low-energy minima in the energy landscape (Holland, 1975; Godberg, 1989). We then examine these deep minima in some detail.

Most applications of GA carried out at the atomic level have involved application to and optimisation of a range of nanoclusters (Deavon and Ho, 1995; Johnston, 2003; Chen et al., 2007; Ferrando et al., 2008). Nevertheless there are an increasing number of successful applications to other materials science problems, such as the prediction of crystal structures (Woodley et al., 1999; Woodley et al., 2008, Oganov et al., 2006, Woodley, 2009), and modelling reconstruction and construction of surfaces (Chuang et al., 2004) and grain boundaries (Zhang et al., 2009; Chua et al., 2010) as well as prediction of ordering in disordered alloys (Johnston et al., 2003; Smith, 1992; Mohn and Kob, 2009; Mohn and Kob, 2011), and searching for alloys with desired physical properties (Dudiy and Zunger 2006). Only a very few examples have demonstrated how GAs can be used to understand ordering patterns or local structure in grossly disordered ceramics. Mohn and Stølen (2005) used a GA to map low energy minima for a binary oxide solid solution but their simulation box was restricted to just 64 ions. Often much larger cells are needed to model gross non-stoichiometry in ceramics (Taylor et al., 1997; Todorov et al., 2003, Bakken et al., 2003). Here we extend the earlier study of Mohn and Stølen (2005) to the more complex carbonate system.

The $\text{CaCO}_3\text{-MgCO}_3$ solid solution itself is one of the most well-examined solid solutions in mineralogy. The essential features are two asymmetric miscibility gaps separated by a narrow stability field for the dolomite (50:50) composition. Calorimetric studies (Navrotsky and Capobianco, 1987; Chai et al., 1995) yield a negative enthalpy of formation for ordered dolomite relative to the end-members MgCO_3 and CaCO_3 . In contrast the enthalpy of formation of a disordered solid solution with the same 50:50 composition is positive (Burton and Kikuchi, 1984; Burton, 1987). Theoretical work has been substantial (Burton and Kikuchi, 1984; Burton, 1987; Davidson, 1994, Burton and Van de Walle, 2003; Purton et

al.,2006) and is consistent with experiment. For example in ref (Vinograd et al., 2007) the fully optimised energies of a large set of randomly varied structures was used to parameterise a cluster expansion of twelve pair-wise effective interactions to obtain the activity-composition relations and a phase diagram in good agreement with experiment.

In the next section we discuss the theoretical methods and the genetic algorithms. Results follow. We then examine the energy landscape in detail, concentrating on the link between the enthalpies of formation and structures of individual configurations. Of particular interest is the local environment of individual Ca^{2+} and Mg^{2+} ions, and how the carbonate ions adjust to accommodate cation neighbours with very different sizes, leading to the observed stability of the ordered dolomite structure. Some brief remarks about extracting thermodynamic properties for such systems from the energy landscape follow, and we also consider the consequences of the form of the energy landscape for the kinetics of transitions between different orderings and hence the difficulty in preparation of ordered dolomite (the so-called “dolomite” problem or paradox).

THEORETICAL METHODS

(a) Energy minimisation using interatomic potentials

For the structural optimisations within the GA algorithm and also for the molecular calculations we have used the set of interatomic shell-model potentials and atomic charges as Fisler et al., (2000). Energy minimisations involved full structural optimisation (Taylor et al., 1997, Taylor et al., 1998) in the static limit (thus ignoring vibrational contributions) of all lattice parameters and atomic positions with *no* symmetry constraints, and were carried out with the GULP code (Gale, 1997). Calculated lattice parameters for MgCO_3 , CaCO_3 and ordered dolomite are in good agreement with experimental values (Table 1).

128

129 Since preliminary runs indicated that inclusion of the shell model in this potential set only
130 affected absolute energies (and then only slightly) but not the relative energies of different
131 arrangements, shells were omitted from the GA runs described below. Shells were included
132 for the molecular mechanics studies described in the final sections of this paper. In addition
133 since we found that in a few high-energy configurations minimisations failed because the
134 carbonate ion became non-planar we increased the four body O-C-O-O torsion constant to
135 0.7510 eV; again this has no effect on any relative energies.

136

137 (b) *Ab initio* optimisations

138 For selected very low and very high energy arrangements we also carried out structural
139 optimisations using the *ab initio* all-electron periodic Hartree-Fock method, as implemented
140 in the CRYSTAL09 code (Dovesi et al., 2005) Previously published basis sets were used
141 (Catti et al., 1991; McCarthy and Harrison, 1994; Catti et al., 1994; Towler et al., 1994) with
142 a Monkhorst-Pack *k*-point grid of $8 \times 8 \times 8$. Once again, no symmetry constraints were
143 applied and calculated lattice parameters are in good agreement with experiment (Table 1).

144

145 (c) Genetic Algorithms

146 The GA used here consists of four steps:

147 1 *Setting up an initial population (of arrangements/configurations)* An initial
148 population is selected at random. Typical population sizes used in GA studies
149 range from a few hundred to several thousand members; here the initial population
150 is 1000. Each member of the initial population is generated by distributing 48 Ca
151 and 48 Mg ions at random over the cation positions in a hexagonal supercell
152 containing, in total, 480 atoms and six cation layers. The energy of each
153 configuration is calculated by a static energy minimisation with full structural
154 relaxation (optimisation) of all basis atom positions and unit cell parameters. We

shall see later that while it is compute-time-intensive this full optimisation is crucial.

- 2 *Selection* Parents with high fitness are preferentially selected using an appropriate scheme. We use a Boltzmann selection where two parents are chosen with a Boltzmann probability, given by $\exp(-\Delta E/kT)$ where ΔE is the energy difference of a configuration relative to the lowest energy so far found and, after a number of test calculations, the value of T was set to 7000 K. The temperature controls the amount of selection i.e. the probability of choosing two parents.

- 3 *3. Mating* We use a real-space crossover where slices and small clusters from the Parent 1 structure are randomly selected and combined with the complementary structure of Parent 2, maintaining the correct composition. We also use a uniform crossover where a random set of cations are selected from Parent 1 and the complementary set of cations is selected from Parent 2, with the constraint that the child has the correct composition. Far less bonds are broken when a real-space crossover is applied compared to that of a binary uniform crossover and the child inherits more local structural information from its parents. On the contrary, a uniform crossover ensures more diversity in the population since no such structural constraints are imposed on the crossover. Results below are reported using a uniform crossover but we compare with calculations carried out using a real-space crossover.

- 4 *4. A full structural optimisation of the child structure is completed, as described above, and the child structure is then added to the population if it has a lower energy than the worst (highest-energy) member in the population, which is itself removed. An important modification that avoids the slow convergence associated with “conventional” GA is the incorporation of the symmetry of the underlying lattice within the GA operators using a randomly chosen symmetry crossover operation, i.e., we simply replace the child with one which is symmetrically equivalent (Mohn and Kob, 2009). This drastically increases the diversity of the population since different symmetrically equivalent regions on the energy landscape are explored simultaneously as explained in detail by Mohn and Kob (2009). In this work only translational symmetry operations along the c -axis were used which creates sufficient diversity to locate the global minima configuration.*

- 5 *Mutation* This involves the exchange of a pair of different cations, chosen at random, in the child. In this work the mutation probability is 0.1.

Steps 2) - 4) are repeated until there are no further changes in the lowest energy arrangements.

It is straightforward to adapt this procedure to search for high-energy rather than low-energy arrangements, provided the method used for the energy calculation is accurate for the different interatomic distances often sampled in such arrangements. In this case, Step 2 is modified by replacing ΔE with $-\Delta E$ so that the members in each population with high energies are selected to mate preferentially, and a child is only added to the population if it has higher energy than the lowest-energy member.

RESULTS

GA

Figure 1 shows the evolution of the lowest-energy member of the population as a function of the number of generations, N . The three dotted lines show the progress of typical calculations each starting from a different initial population of 1000 randomly-generated structures. The red curve is the average lowest-energy after N generations of the results obtained from 100 such initial populations. The crystal structure is shown at three different stages of one of the calculations, showing the emergence of intact layers along the c -axis containing single cation types and finally the emergence of ordered dolomite as the lowest-energy structure. This global minimum structure contains alternating layers of Mg and Ca ions along the c -axis, and each individual cation layer contains only one type of ion. 94% of all runs reached the global minimum within 20000 generations and 80% of all runs within 10000 generations.

The speed at which the GA can find orderings very similar in structure and energy to ordered dolomite with layers of one only cation type along the c -axis is striking. Searching for low-

energy structures is generally a very challenging problem for unit cells with, as here, at least several hundred atoms. There are more than 6.4×10^{27} arrangements of the cations for the unit cell used in this work, which contains a total of 96 cations of which 50% are Mg. The fact that the global minimum structure can be found quickly and reliably demonstrates the effectiveness of GA methods for searching for low-energy structures in systems of this type.

The benefit of including the symmetry crossover operations is evident. The global minimum structure was only found in 3% of all runs after 20000 steps, in contrast to 94% when the symmetry operations were used. When a real-space crossover is applied instead of an uniform crossover the successrate was slightly lower because the diversity in the population is larger when an uniform crossover is used. However, when using a real space crossover the convergence to the global minima was slightly faster since fewer bonds are broken after crossover since the child inherits more structural information from its parents (the crossover is more efficient). It is worth bearing in mind that when full relaxations of the structures generated in each generation are not included, the GA algorithm fails to find the ordered dolomite structure. This emphasises the necessity of including full relaxations of all the atom positions and unit cell parameters, despite the considerable extra expense in computer time. The reasons at the atomic level for this will become clearer in the next section.

The GA algorithm is readily reversed to generate states which are high, rather than low, in energy. There is no one global maximum, but rather a large number of structures close in energy in which each layer along the *c*-axis contains both Mg and Ca ions.

237

238 **Structure analysis**

239

240 We have probed the important features of low- and high-energy structures through molecular
 241 mechanics calculations, carried out with full structural optimisations and interatomic
 242 potentials as described earlier. For comparison, geometry optimisations and total energies
 243 were also calculated using the periodic *ab initio* Hartree-Fock method. We wish to examine
 244 the structural changes that accompany the differences in energy between different cation
 245 orderings and ultimately give rise to the strong non-ideality of the $\text{CaCO}_3\text{-MgCO}_3$ solid
 246 solution. Four different orderings of $\text{Mg}_{0.5}\text{Ca}_{0.5}\text{CO}_3$ have been investigated (Figure 2),
 247 selected as those of particular interest based on the GA results.

248 For these studies we have used a unit cell containing twelve cations (six cation layers with
 249 two cations per layer). Energies (in the static limit) are reported as the energy of formation of
 250 the mixed system relative to that of the pure end-members:

251

$$252 \quad E_{\text{formation}} = E(\text{Mg}_{0.5}\text{Ca}_{0.5}\text{CO}_3) - \frac{1}{2}E(\text{MgCO}_3) - \frac{1}{2}E(\text{CaCO}_3)$$

253

254 The energies of the four cation orderings are given in Table 2. While the numbers differ
 255 quantitatively, all the trends are the same with both molecular mechanics and Hartree-Fock
 256 methods. The Hartree-Fock formation energies are all slightly less negative (or more positive)
 257 than those calculated with the interatomic potentials. For comparison, Chan and Zungers
 258 formation energy (Chan and Zunger, 2009) for ordered dolomite calculated using density
 259 functional theory in the generalised gradient approximation is -39 meV per cation pair, while

the experimentally determined value is approximately -100 meV per cation pair (Navrotsky, 1987).

262

263

264 DISCUSSION

265 The results from the molecular mechanics calculations show that, in the structures in
266 which each layer contains cations of only a single element (Orderings 1, 2 and 3), the
267 carbonate groups lie flat between the cation layers (Figure 2a-c), as they do in the end-
268 members CaCO_3 and MgCO_3 .

269 In CaCO_3 and MgCO_3 , the carbonate group is oriented to give identical Ca-O/Mg-O
270 distances to cations in the layers both above and below the anion layer (Figure 2a and b). In
271 ordered dolomite, clearly the cation-O distances should not all be identical, but Ca-O
272 distances should be significantly longer than Mg-O. If the same orientation of the carbonate
273 groups were maintained in ordered dolomite as in CaCO_3 and MgCO_3 , the only degree of
274 freedom allowing optimisation of the Mg-O and Ca-O distances would be the interlayer
275 spacing. However, an additional degree of freedom is obtained through rotation of the
276 carbonate group. Thus, when there are alternating layers of Ca and Mg ions along the *c*-axis
277 (as in ordered dolomite, Ordering 1, Figure 2a), the carbonate group rotates within the *ab*-
278 plane so that the oxygen atoms are closer to the Mg ions in one neighboring cation layer and
279 further from the calcium ions in the other neighbouring cation layer (Figure 3c). Hence, the
280 rotation of the carbonate groups is crucial since it allows both the Mg-O and Ca-O distances
281 to be optimised without introducing significant strain into the C-O bonds. The rotation of the
282 carbonate group in ordered dolomite relative to that in CaCO_3 (and MgCO_3) is shown in
283 Figure 3d.

284 Like ordered dolomite, Orderings 2 and 3 also have intact Mg^{2+} and Ca^{2+} layers

(Figures 2b and c), but are higher in energy. In Ordering 3, for which Table 2 shows that the formation energy is positive, there are three adjacent layers of cations of the same element (Figure 2b). The lattice parameters are intermediate between those of CaCO_3 and MgCO_3 , so the CaCO_3 layers are compressed in the *ab*-plane relative to pure calcite, while the MgCO_3 layers are stretched relative to pure magnesite (Figure 3e). The Ca-O distances are thus shorter than those in the end-member CaCO_3 , while Mg-O distances are longer. Some strain is also found in the C-O bonds within the carbonate groups, with these bonds being significantly compressed in anion layers adjacent to Ca (Figure 3e) and stretched in anion layers adjacent to Mg.

Except at the interfaces between the Ca and Mg layers, these strains in the bond lengths cannot be relieved by rotations of the carbonate groups, in contrast to ordered dolomite. In these structures with two or more adjacent layers with cations of the same element (e.g. Orderings 2 and 3, Figure 2b and c), no advantage can be gained through rotation of the carbonate groups. For example, rotation to decrease the Mg-O distances in one direction (e.g. to the layer below) would increase them in the other direction (e.g. to the layer above). Hence, the carbonate groups between two layers of the same cation remain in the same orientation as in pure magnesite and calcite; the bond lengths are compromised because the lattice parameters are in between those of the end-members and there are insufficient degrees of freedom for bond length optimisation.

Thus, rotation of the carbonate groups to optimise bond lengths is *only* effective when the anion layer has Mg ions in one neighbouring layer and Ca ions in the other, and so the lowest energy ordering for $\text{Mg}_{0.5}\text{Ca}_{0.5}\text{CO}_3$ is that with alternating layers of Ca and Mg ions. Ordering 2 has a formation energy intermediate between those of ordered dolomite (Ordering 1) and Ordering 3 (Table 2); the number of anion layers with Mg ions in one neighbouring layer and calcium ions in the other is also intermediate between Ordering 1 and Ordering 3.

310 This again indicates the energetic favorability of the cation species alternating between
 311 layers, so that the carbonate groups can rotate to achieve optimal O-Ca and O-Mg
 312 separations.

313 In Ordering 4, in which all layers contain a mixture of Mg^{2+} and Ca^{2+} (Figure 2d), the
 314 carbonate groups are distorted and no longer lie flat between the cation layers. There is
 315 considerable variation in C-O bond lengths in this particular ordering and O-C-O bond angles
 316 range from 117.2° to 122.5° . In contrast, in the ordered dolomite structure, all C-O bond
 317 lengths are the same and all O-C-O bond angles are 120.0° (as they are in the CaCO_3 and
 318 MgCO_3 end-members). Similarly, in the ordered dolomite structure, all Mg-O and Ca-O
 319 distances are the same, whereas there is significant variation in these distances for Ordering
 320 4. Several different orderings with layers containing a mixture of Mg^{2+} and Ca^{2+} have been
 321 investigated in addition to Ordering 4; all have positive energies of formation, with Ordering
 322 4 the maximum. This indicates that layers containing a mixture of cations are unfavourable,
 323 consistent with the GA results. Chan and Zunger (2009) also found that a random cation
 324 distribution is high in energy and accompanied by carbonate ion distortions and strain. All the
 325 same qualitative conclusions and trends are also evident in the results from the Hartree-Fock
 326 calculations.

327

328 Ordered dolomite has a small but negative enthalpy of mixing (Table 2). The rotation
 329 of the carbonate group also allows more volume-efficient stacking of the layers than in either
 330 MgCO_3 or CaCO_3 end-members. The separation along the *c*-axis of a layer of Mg^{2+} ions and
 331 the neighbouring anion layer is 1.249 \AA in MgCO_3 and 1.228 \AA in ordered dolomite; the
 332 separation in the *c*-axis direction of a layer of Ca^{2+} ions and the neighbouring anion layer is
 333 1.421 \AA in CaCO_3 and 1.409 \AA in ordered dolomite. Thus the volume of ordered dolomite is
 334 $\sim 1 \%$ smaller than the average volume of MgCO_3 and CaCO_3 and at fixed temperature the

thermodynamic stability of ordered dolomite with respect to the end-members increases with increasing pressure.

The nature of the low energy excited states as revealed by both the GA and the molecular mechanics calculations also provides some insight into the difficulty of formation of ordered dolomite (Deelman, 1999) If we “funnel” down in energy we end up in local minima with structures in which each cation layer contains ions of only one element but the Ca^{2+} and Mg^{2+} are not alternating; there are substantial kinetic barriers in proceeding further to the global minimum, since this would require interchange of cations between layers, producing intermediate high-energy structures with a mixture of Ca and Mg cations in the layers. The activation energies for such cation interchanges are large.

We end with a few remarks about the consequences of non-ideality for calculation of thermodynamic properties, in particular by configurational averaging (often referred to as basin sampling). In principle a solid solution can assume any state, i.e., each atom can be at any position and each will have a different probability. However, the only states of practical importance away from the melting point lie at the bottom of K local minima in the energy landscape, so the thermodynamic averaging is carried out over results from a set of optimisations of different cation arrangements within a given supercell. The configurational averaging approach to solid solutions commonly uses the isobaric-isothermal (NPT) ensemble, so, for example, the enthalpy of the solid solution H is given by;

$$H = \frac{\sum_k H_k \exp(-\beta G_k)}{\sum_k \exp(-\beta G_k)} \quad (1)$$

where G_k and H_k are the free energy and enthalpy of each local minimum respectively. All vibrational entropy terms are usually neglected so G_k is replaced by H_k calculated in the static limit.

For other than the smallest supercells it is impractical to sum over all K configurations and so both summations in Eq. (1) are restricted to K' configurations chosen at random. In previous work on non-ideal solid solutions showing much smaller deviations from non-ideality than magnesite-calcite, we demonstrated convergence with a manageable value of K' configurations, chosen at random. For a 32-atom supercell of composition 50% MgO /50% MnO convergence of the formation enthalpy of the solid solution to 0.04 kJ mol^{-1} is typically obtained with only ≈ 150 out of a total of 12,870 configurations (Allan et al., 2001).

This procedure requires adaption for very strongly non-ideal systems such as MgCO_3 - CaCO_3 . For so-called “disordered” dolomite, a random selection of configurations can be used, excluding any with negative heats of formation. In any case, in any reasonably sized simulation cell the chance of selection of an ordering with a negative heat of formation is extremely small due to the small weightings of the highly-ordered states which are the only states with such exothermic heats of formation. For *ordered* dolomite a modified procedure is needed. A working procedure is to select a random set of configurations including the lowest energy state and any state of energy within kT of the ground state as indicated by the GA. In practice it is more accurate not to include the weightings of the individual states when using equation (1) with a small number of configurations; this apparently counterintuitive conclusion arises because of the two summations in this equation. Including only a few states fully weighted tends to underestimate the partition function in the denominator and so overestimates the final result. In our previous work (Purton et al., 2004) we did not follow such a procedure and consequently our results for the enthalpy of mixing were overestimated. The modified method gives values both for the enthalpy of formation for dolomite and its temperature variation very close to those obtained using the Cluster Variation Method in (Burton and Van de Walle, 2003).

383 CONCLUSIONS

384 In this paper we have shown that a GA incorporating symmetry is a particularly
385 computationally efficient method of establishing non-ideality and any preferential ordering in
386 a solid solution. Combined with configurational averaging (basin sampling) it thus provides a
387 very powerful tool for modelling solid solutions and non-stoichiometry in general. It readily
388 provides information as to low-lying and higher excited states. Extension to high pressures
389 would be straightforward. Here the generated energy landscape provides atomistic insights
390 into why dolomite forms – the rotation of the carbonate ions between adjacent layers of Ca
391 and Mg relieves strain – and also into the dolomite problem. The presence of a polyatomic
392 ion is thus crucial for the formation of an ordered mixed phase – the same reduction of strain
393 is impossible in the MgO-CaO binary solution despite the similarity of the rock salt and
394 calcite structures. The understanding of local order in substitutionally disordered materials is
395 important in fields as diverse as the development of new materials with improved mechanical
396 or electrical properties and the understanding of fundamental geochemical processes in the
397 deep Earth, and we hope the techniques and the encouraging results presented in this paper
398 will assist in such investigations.

399

400 ACKNOWLEDGEMENTS

401 NLA is grateful for valuable discussions with Victor Vinograd which prompted this work.
402 This work was, in part, performed on the Abel Cluster, owned by the University of Oslo and
403 the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the
404 Department for Research Computing at the University of Oslo IT-department. CM
405 acknowledges support from the Research Council of Norway through its Centres of
406 Excellence funding scheme, project number 223272.

407

408

409 **REFERENCES**

410

411 Allan N.L., Barrera G.D., Fracchia R.M., Lavrentiev M. Yu., Taylor M.B., Todorov I.T. and Purton
 412 J.A. (2001) Free energy of solid solutions and phase diagrams via quasiharmonic lattice dynamics.
 413 *Phys. Rev. B* **63**, 094203.

414

415 Althoff P.L. (1977) *American Mineral.* **62**, 772-783.

416

417 Burton B.P. and Kikuchi R. (1984) Thermodynamic analysis of the system $\text{CaCO}_3\text{--MgCO}_3$ in the
 418 tetrahedron approximation. *American Mineral.* **69**, 165-175.

419

420 Burton B.P. and Van de Walle A. (2003) First principles based calculations of the $\text{CaCO}_3\text{--MgCO}_3$
 421 subsolidus phase diagrams. *Phys. Chem. Miner.* **30**, 88–97.

422

423 Catti M., Dovesi R., Pavese A. and Saunders V.R. (1991). Elastic constants and electronic structure of
 424 fluorite (CaF_2): an ab initio Hartree-Fock study, *J. Phys. Condens. Matter* **3**, 4151-4164.

425

426 Catti M., Pavese A., Dovesi R. and Saunders V.R. (1993) Static lattice and electron properties of
 427 MgCO_3 (Magnesite) calculated by ab initio periodic Hartree-Fock methods. *Phys. Rev. B* **47**, 9189-
 428 9198.

429

430 Chai L., Navrotsky A. and Reeder R.J. (1995) Energetics of calcium-rich dolomite. *Geochim.*
 431 *Cosmochim. Acta* **59**, 939–944; Chai, L. and Navrotsky, A. (1996) Synthesis, characterization, and
 432 energetics of solid solution along the $\text{CaMg}(\text{CO}_3)_2\text{--CaFe}(\text{CO}_3)_2$ join and implication for the stability
 433 of ordered $\text{CaFe}(\text{CO}_3)_2$. *American Mineral.* **81**, 1141–1147; Navrotsky A., Dooley D., Reeder R. and
 434 Brady P. (1999) Calorimetric studies of the energetics of order-disorder in the system $\text{Mg}_x\text{Fe}_{1-x}\text{Ca}(\text{CO}_3)_2$.
 435 *American Mineral.* **84**, 1622–1626.

436

437 Chan J.A. and Zunger A. (2009) II-VI oxides phase separate whereas the corresponding carbonates
 438 order: The stabilizing role of anionic groups. *Phys. Rev. B* **80**, 165201

439

440 Chen F.Y., Curley B.C., Rossi, G., Johnston R.L. (2007) Structure, melting, and thermal stability of 55
 441 atom Ag–Au nanoalloys. *J. Phys. Chem.* **C111**, 9157-9165.

442

443 Chua A. L.-S., Benedek N.A., Chen L., Finnis M.W. and Sutton A.P. (2010) A genetic algorithm for
 444 predicting the structures of interfaces in multicomponent systems *Nature Materials* **9**, 418–422.

445

446 Chuang F.C., Ciobanu C.V., Shenoy V.B., Wang C.Z. and Ho K.M. (2004) Finding the reconstructions
 447 of semiconductor surfaces via a genetic algorithm. *Surf. Sci.*, **573**, L375-L381.

448

449 Davidson, P.M. (1994) Ternary iron, magnesium, calcium carbonates: a thermodynamic model for
 450 dolomite as an ordered derivative of calcite structure solutions. *American Mineral.* **79**, 332–339

451

452 Deaven D.M. and Ho K.M. (1995) Molecular-Geometry Optimization with a genetic algorithm *Phys.*
 453 *Rev. Lett.*, **75**, 288-291.

454

455 Deelman J.C. (1999) Low temperature nucleation of magnesite and dolomite, *Neues Jahrbuch für*
 456 *Mineralogie, Monatshefte* **7**, 289-302.

457

458 Dovesi R., Orlando R., Civalleri B., Roetti C., Saunders V.R. and Zicovich-Wilson C.M. Z.
 459 *Kristallogr.* **220**, 571-573 (2005); Dovesi R., Saunders V.R., Roetti C., Orlando R., Zicovich-Wilson

- C.M., Pascale F., Civalleri B., Doll K., Harrison N.M., Bush I.J., D'Arco P. and Llunell M. (2009) CRYSTAL09 User's Manual, University of Torino, Torino.
- Dudiy S.V. and Zunger A. (2006) Searching for Alloy Configurations with Target Physical Properties: Impurity Design via a Genetic Algorithm Inverse Band Structure Approach. *Phys. Rev. Lett.* **97**, 046401.
- Ferrando R., Fortunelli A., and Johnston R.L. (2008) Searching for the optimum structures of alloy nanoclusters. *Phys. Chem. Chem. Phys.* **10**, 640-649.
- Fisler D.K., Gale J.D. and Cygan R.T (2000), *American Mineral.* **85**, 217-224.
- Gale J.D. (1997) The General Utility Lattice Program GULP - a computer program for the symmetry adapted simulation of solids. *J. Chem. Soc., Faraday Trans.* **93**, 629-637 ; Gale J.D. and Rohl A.L. (2003) *Mol. Simul.* **29**, 291-34; J.D. Gale (2005) GULP: Capabilities and prospects, *Z. Krist.* **220**, 552-554.
- Goldberg D.E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison Wesley: Reading,
- Holland J.H. (1975) *Adaption in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor.
- Johnston R.L. (2003) Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries. *Dalton. Trans.*, **22**, 4193-4207.
- McCarthy M.I. and Harrison N.M. (1994) Ab Initio determination of the bulk properties of MgO. *Phys. Rev. B* **49**, 8574-8582.
- Mohn C.E. and Kob W. (2009) A genetic algorithm for the atomistic design and global optimisation of substitutionally disordered materials. *Comp. Mater. Sci.* **45**, 111-117.
- Mohn C.E. and Kob W. (2015) Predicting complex mineral structures using genetic algorithms *J. Phys: Condens. Matter* **47**, 425201
- Mohn C.E., Stølen S. and Kob W. (2011) Predicting the structure of alloys using genetic algorithms. *Materials and Manufacturing Processes* **26**, 348–353
- Mohn C.E., Lavrentiev M.Y., Allan N.L., Bakken E., Stølen S. (2005) *Size mismatch effects in oxide solid solutions using Monte Carlo and configurational averaging*, **7**, 1127-1135
- Mohn C.E. and Stølen S. (2005) Genetic mapping of the distribution of minima on the potential energy surface of disordered systems *J. Chem. Phys.*, **123**, 114104.
- Navrotsky A. and Capobianco C. (1987) Enthalpies of formation of dolomite and of magnesian calcites. *American Mineral.* **72**, 782–787.
- Navrotsky A. (1987) Models of crystalline solutions. In: Reeder, R.J. (Ed.), *Reviews in Mineralogy*, Vol. 17. Mineralogical Society of America, Washington DC.
- Oganov A. R. and Glass, C. W. (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704; Oganov A.R., Chen J., Gatti C., Ma Y.-Z., Ma Y.-M., Glass C.W., Liu Z., Yu T., Kurakevych O.O., Solozhenko V.L. (2009) Ionic high-pressure form of elemental boron. *Nature* **457**, 863-867; Lyakhov A.O., Oganov A.R. and Valle, M. (2010) How to predict very large and complex crystal structures, *Computer Physics Communications*

- 181**, 1623-1632, Oganov A.R., Lyakhov A.O. and Valle M. (2011). How evolutionary crystal structure prediction works - and why. *Acc. Chem. Res.* **44**, 227-23; Zhu Q., Oganov A.R., Glass C.W. and Stokes H.T. (2012). Structure prediction for molecular crystals using evolutionary algorithms: methodology and applications. *Acta Cryst.* **B68**, 215-226; Oganov A.R., Ma Y., Lyakhov A.O., Valle M. and Gatti C. (2010). Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. *Rev. Mineral. Geochem.* **71**, 271-298; Lyakhov A.O., Oganov A.R., Stokes H.T. and Zhu Q. (2013). New developments in evolutionary structure prediction algorithm USPEX. *Comp. Phys. Comm.* **184**, 1172-1182
- Purton J.A., Allan N.L., Lavrentiev M.Yu., Todorov I.T. and Freeman C.L. (2006) Computer simulation of mineral solid solutions. *Chemical Geology* **225**, 176-188
- Purton J.A., Blundy J.D., Taylor M.B., Barrera G.D. and Allan N.L. (1998) Hybrid Monte Carlo and lattice dynamics simulations: the enthalpy of mixing of binary oxides, *Chem. Commun.* 627-628.
- Purton J.A., Barrera G.D., Allan N.L. and Blundy J.D. (1998) Monte Carlo and hybrid Monte Carlo/molecular dynamics approaches to order-disorder in alloys, oxides and silicates. *J. Phys. Chem.* **B102**, 5202-5207; Purton J.A., Lavrentiev M. Yu. and Allan N.L. (2007) Monte Carlo simulation of GaN/AlN and AlN/InN mixtures, *Mat. Chem. Phys.* **105** 179-184; Purton J.A., Parker S.C. and Allan N.L. (2013) Monte Carlo simulation and free energies of mixed oxide nanoparticles. *Phys. Chem., Chem. Phys.* **15**, 6219-6225.
- All ionic radii are for six-fold coordination and taken from Shannon R.D. (1976), Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystall.* **A32**, 751-767.
- Smith R.W. (1992) Energy minimization of binary alloy models via genetic algorithms. *Comp. Phys. Commun.* **71**, 134-146.
- Taylor M.B., Barrera G.D., Allan N.L., Barron T.H.K. and Mackrodt W.C. (1997) The free energy of formation of defects in polar solids. *Faraday Discuss.* **106**, 377-387.
- Taylor M.B., Barrera G.D., Allan N.L. and Barron T.H.K. (1997) Free energy derivatives and structure optimisation within quasiharmonic lattice dynamics, *Phys. Rev.* **B56**, 14380-14390.
- Taylor M.B., Barrera G.D., Allan N.L., Barron T.H.K. and Mackrodt W.C. (1998) SHELL – a code for lattice dynamics and structure optimisation of ionic crystals, *Computer Phys. Commun.* **109**, 135-143.
- Todorov I.T., Allan N.L., Lavrentiev M. Yu., Freeman C.L., Mohn C.E. and Purton J.A. (2004) Computer simulation of mineral solid solutions *J. Phys.: Condens. Matter* **16**, S2751-S2770
- Towler M.D., Allan, N.L., Harrison N.M., Saunders V.R., Mackrodt W.C. and Aprà E. (1994) Ab initio Hartree-Fock study of MnO and NiO. *Phys. Rev. B* **50**, 5041-5054.
- Vinograd V.L., Burton B.P., Gale J.D., Allan N.L. and Winkler B. (2007) Activity-composition relations in the system CaCO₃-MgCO₃ predicted from static structure energy calculations and Monte Carlo simulations. *Geochimica et Cosmochimica Acta* **71**, 974-983.
- For example, van Westrenen W., Allan N.L., Blundy J.D., Lavrentiev M. Yu., Lucas B.R. and Purton J.A. (2003) Trace element incorporation into pyrope-grossular solid solutions: an atomistic simulation study. *Phys. Chem. Minerals* **30**, 217-229; M.H.F. Sluiter, V. Vinograd, and Y. Kawazoe (2004) Intermixing tendencies in garnets: Pyrope and grossular. *Phys. Rev.* **B70**, 184120; C.L. Freeman, N.L. Allan and W. van Westrenen (2006) Local cation environments in the pyrope-grossular Mg₃Al₂Si₃O₁₂-Ca₃Al₂Si₃O₁₂ garnet solid solution. *Phys. Rev. B* **74** 134203; Lavrentiev, M. Yu, van Westrenen W., Allan N.L., Freeman C.L. and Purton J.A. (2006) Simulation of thermodynamic mixing properties of

- 570 garnet solid solutions at high temperatures and pressures. *Chem. Geol.* **225** 336-346.
- 571
- 572 Woodley S.M. (2009) Structure prediction of ternary oxide subnanoparticles. *Materials and*
- 573 *Manufacturing Processes*, **24**, 255-264.
- 574
- 575 Woodley S.M., Battle P.D., Gale, J.D., Catlow, C.R.A. (1999) The prediction of inorganic crystal
- 576 structures using a genetic algorithm and energy minimisation. *Phys. Chem. Chem. Phys.* **1**, 2535-
- 577 2542.
- 578
- 579 Woodley S.M and Catlow, C.R.A. (2008) Crystal structure prediction from first principles. *Nature*
- 580 *Materials*, **7**, 937-946.
- 581
- 582 Zhang J. and Reeder R.J. (1999) *American Mineral.* **84**, 861-870.
- 583
- 584 Zhang J., Wang C. Z. and Ho, K. M. (2009) Finding the low-energy structures of Si[001] symmetric
- 585 tilted grain boundaries with a genetic algorithm. *Phys. Rev. B* **80**, 174102.

586 Table 1. Calculated and experimental lattice parameters for magnesite, calcite and
587 dolomite.

588

589 Table 2. Optimised volumes and formation energies, in the static limit, calculated using
590 molecular mechanics and (in parentheses) periodic Hartree-Fock theory, of some different
591 cation orderings in $\text{Mg}_{0.5}\text{Ca}_{0.5}\text{CO}_3$. Ordering 1 is ordered dolomite with alternating
592 layers of Ca and Mg along the *c*-axis. Note the large differences in energy between the
593 four orderings.

594 Figure 1. Progression plots (lowest unit cell energy vs. number of generations *N*) of the
595 genetic algorithm for $\text{Mg}_{0.5}\text{Ca}_{0.5}\text{CO}_3$ using a uniform crossover. The energy value is
596 relative to the global energy minimum. The unit cell contained 480 atoms (96 cations).
597 The dotted lines show the progress of the calculations starting from three different initial
598 populations of 1000 randomly-generated structures. The red curve shows the energy
599 obtained from averaging the results at each generation for 100 such starting populations.
600 In the accompanying crystal structures, the *a*-axis points out of the plane of the paper, Mg
601 ions are orange, Ca blue, C grey and oxygen red. The black arrows show the step and the
602 GA run to which these structures relate.

603

604 Figure 2. (a) Ordered dolomite (Ordering 1), (b) Ordering 2, (c) Ordering 3 and (d)
605 Ordering 4. Corresponding formation energies are listed in Table 2. The blue atoms are
606 Ca atoms, the orange atoms Mg, the brown atoms carbon and the red atoms oxygen. The
607 black lines show the unit cell boundaries. The *c*-axis direction is the vertical direction.

608

609 Figure 3. Two cation layers (*ab*-plane) and one anion layer of (a) CaCO_3 , (b) MgCO_3 , (c)
610 ordered dolomite (Ordering 1), viewed along the *c*-axis (cations in the layer below the

611 anion layer are labelled “1”; cations above the anion layer are labelled “2”). (d) Ordered
612 dolomite with the positions of the selected atoms in the end members MgCO_3 and CaCO_3
613 overlaid (pink atoms show the position of the oxygen atoms in MgCO_3 and CaCO_3). Note
614 that, in ordered dolomite, the carbonate group rotates clockwise to move the oxygen
615 atoms closer to Mg ions in the layer above and further from Ca ions in the layer below,
616 hence optimizing both the Mg-O and Ca-O distances. (e) Ordering 3 (the calcium layer
617 shown is the middle of the three adjacent layers, Figure 2b). Blue atoms are Ca, orange
618 atoms are Mg, brown atoms are carbon and red atoms are oxygen. The black lines show
619 the unit cell boundaries. Some important bond lengths are shown; note that the Ca-O and
620 C-O bond lengths in ordered dolomite (c) are closer to those of pure calcite (a) than in
621 Ordering 3 (e), in which these bond lengths are reduced.

	MgCO ₃		CaCO ₃		Dolomite	
	<i>a</i> / Å	<i>c</i> / Å	<i>a</i> / Å	<i>c</i> / Å	<i>a</i> / Å	<i>c</i> / Å
Molecular mechanics	4.679	14.915	4.991	17.057	4.849	15.818
<i>Ab initio</i> Hartree-Fock	4.648	15.092	5.065	17.234	4.842	16.189
Experimental	4.636 ^a	15.021 ^a	4.989 ^a	17.042 ^a	4.807 ^b ,4.803 ^c	16.003 ^b ,15.984 ^c

^a Ref (Chang and Reeder, 1999), ^b ref Taylor, 1997) and ^c ref (Althoff, 1977)

Ordering	Volume / Å ³ per cation pair	$E_{\text{formation}}$ / meV per cation pair
1	107.37 (109.59)	-80 (-52)
2	107.64 (109.91)	-13 (52)
3	107.93 (110.20)	54 (158)
4	107.53 (110.34)	169 (281)





